AFRL-RH-WP-TR-2011-0075

INITIAL EXPERIMENTATION ON AUDIO ANNOTATION USING A DISTRIBUTED VIRTUAL ENVIRONMENT

Robert H. Gilkey
Thomas J. Moore

Wright State University

JANUARY 2011
Interim Report

Distribution A: Approved for public release; distribution unlimited.

AIR FORCE RESEARCH LABORATORY
711 HUMAN PERFORMANCE WING,
HUMAN EFFECTIVENESS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

AFRL-RH-WP-TR-2011-0075 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signed//
Brian D. Simpson
Senior Engineering Research Psychologist
Battlespace Acoustics Branch

//signed//
William T. Nelson
Chief, Battlespace Acoustics Branch
Warfighter Interface Division

//signed//
Michael A. Stropki
Chief, Warfighter Interface Division
Human Effectiveness Directorate
711 Human Performance Wing

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE ***(DD-MM-YY)***
01/01/11

2. REPORT TYPE
Interim

3. DATES COVERED ***(From - To)***
28 July 2009 – 28 October 2010

4. TITLE AND SUBTITLE
INITIAL EXPERIMENTATION ON AUDIO ANNOTATION USING A DISTRIBUTED VIRTUAL ENVIRONMENT

5a. CONTRACT NUMBER
IN-HOUSE

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)
Robert H. Gilkey
Thomas J. Moore

5d. PROJECT NUMBER
7184

5e. TASK NUMBER
16

5f. WORK UNIT NUMBER
71841621

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Wright State University
3640 Colonel Glenn Highway
Dayton, OH 45435

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command
Air Force Research Laboratory
711 Human Performance Wing
Human Effectiveness Directorate
Warfighter Interface Division
Battlespace Acoustics Branch
Wright-Patterson Air Force Base, OH 45433

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
711HPW/RHCB

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-RH-WP-TR-2011-0075

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution A: Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
88ABW Cleared 09/23/2011; 88ABW-2011-5172. Report contains color.

14. ABSTRACT: Today's battlespace has shifted from one of traditional warfare to non-conventional warfare. In addition to airborne operations, the USAF now employs personnel fighting as dismounted operators in dynamic and dangerous urban environments. New sensor capabilities will support the collection of information from multiple levels of the battlespace. Howevever, for that information to support effective decision making, it must be disseminated in a rapid, unambiguous manner and enable operators to remain 'heads-up' and 'eyes-out' to safely complete their mission. This efforts is part of a larger program of research to develop displays to increase mission effectiveness for dismounted personnel in ground operations. Here, we focus on the design and development of wearable 3D audio displays, and Audio Annotation in particular, which were tested in a virtual environment portraying an operationally relevant scenario in an urban setting. Results suggest that an audio display can lead to comparable or superior performance on a simulated search and rescue task relative to a visual display. Additional benefits of audio annotation relative to a mono display were not seen here. This has been attributed to the apparent ceiling performance achieved by subjects due to insufficient overall task loading. Future experiments will address these issues.

15. SUBJECT TERMS
3D Audio, Audio Displays, Audio Annotation, Virtual Environments, Urban Warfare, Search and Rescue

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
a. REPORT	b. ABSTRACT	c. THIS PAGE			Brian D. Simpson
Unclassified	Unclassified	Unclassified	SAR	17	19b. TELEPHONE NUMBER *(Include Area Code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18

Table of Contents

List of Figures

Figure **Page**

List of Tables

Table **Page**

THIS PAGE IS INTENTIONALLY LEFT BLANK.

1.0 SUMMARY

The present interim report describes research conducted by Wright State University personnel in support of an Air Force Research Laboratory research program that has as its goal the development of displays that provide mission critical information to combat personnel in an intuitive and timely manner. In particular this research investigates what advantage, if any, a spatialized audio display provides combat personnel over other commonly employed information displays (e.g., standard audio displays, visual displays). The synthetic task used to evaluate the effectiveness of the displays is a "search-and-rescue" scenario. Pairs of subjects performed as a distributed team (i.e., each member of the pair was physically located in a different immersive virtual environment). One member of the pair performed as a ground-based para-rescue jumper (PJ) searching for a downed pilot while navigating through a virtual urban environment. The other member of the team performed as a helicopter based "Spotter" whose task was to provide information via the displays being evaluated to guide the PJ through the virtual urban environment to the location of the downed pilot as quickly as possible, while also indicating the presence and location of enemy personnel.

A number of dependent measures were statistically analyzed to evaluate the relative performance of the displays being evaluated. While there were some statistically significant differences in performance between the displays being evaluated, these effects were relatively small. Overall, there were no major differences in performance attributable to the type of display evaluated. Several reasons for the lack of major effects and lessons learned for future research are discussed.

2.0 INTRODUCTION

This contract is part of a larger program of research aimed at increasing mission effectiveness and soldier survivability in ground operations. In particular, we have considered the role of the Pararescue Jumper (PJ) in the Combat Search and Rescue (CSAR) mission. Our focus has been on the design and evaluation of wearable 3D audio displays as a means to efficiently provide mission critical information in a manner that minimally interferes with the processing of other displays and environmental information.

The work reported here builds on a previous subcontract from Ball Aerospace in which synthetic scenarios were developed for future experimental studies. In turn, the Ball contract built on previous contracts and subcontracts from AFRL/RH and daytaOhio. The current contract developed and executed initial experimental protocols to evaluate audio annotation (described below).

Warfighters in general, and ground soldiers in specific, receive wide-ranging information from a variety of sources. This information is received both directly from the environment around them and indirectly through various maps, visual displays, communication channels, etc. It is important to design display systems that draw attention to the most critical information and present that information in a manner that can be rapidly understood. In addition, an ideal display would minimally interfere with the perception and processing of other sources of information. The location of a person, object, or event of interest is one piece of information that is often essential, but can be difficult for team members to communicate. Consider, for example, the situation depicted in the left panel of Figure 1. Here, a team member with an aerial view must rapidly convey the location of a sniper to a column of vehicles and soldiers as they progress through a city. Note that in this situation the person with the aerial view cannot easily use terms like "left" or "right" or clock angles, because

the sniper is in a different relative position to each person in the column. Although it would be possible to highlight the location on some form of heads-down map display, it is unlikely that each person would have access to such a display and those that do may need to keep their heads up to deal with the surrounding environment. So, as depicted in the left panel of Figure 1, a lengthy description must be given (in this case "Sniper on B Street, on the roof of the roof of the building to the right of the 3rd vehicle in the southbound caravan"), which may be difficult to interpret (e.g., some team members may not be on B Street yet and/or may not be able to see the 3rd vehicle in the caravan). The right panel of Figure 1 depicts the same situation, but using audio annotation. Here the team member with the aerial view uses a laser range finder to mark the location of the sniper and simply says "Sniper over here." The verbal message, along with the GPS coordinates of the sniper, are sent to each team member. Wearable computers are used to spatialize the sound relative to the GPS location of each particular team member. The team members then orient to the sound as they would to any other threat.

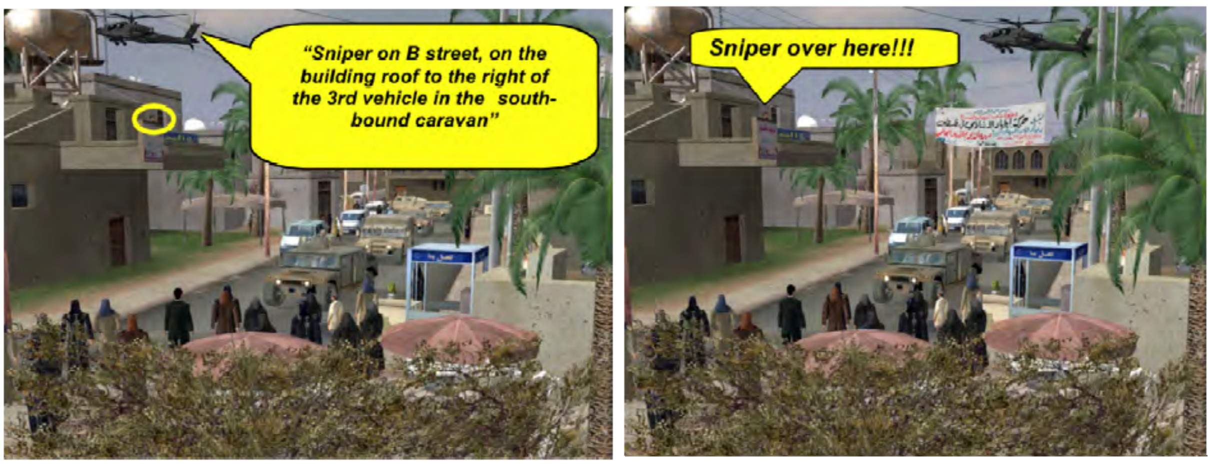

Figure 1. Communications to pinpoint the location of a sniper to other team members: Left panel -- using a traditional communication channel (i.e., not spatialized), and right panel -- using audio annotation (i.e., the communication channel is spatialized, separately for each team member, so that the message sounds as if it is coming from the location of the sniper (adapted from Ephrem, Finomore, Gilkey, Newton, Romigh, Simpson, Brungart, and Cowgill, 2008).

3.0 DEVELOPMENT

To support this work a variety of hardware and software developments were implemented. With funds from the Ohio Department of Development (received through daytaOhio) the CAVE® in the Virtual Environment Research, Interactive Technology, and Simulation (VERITAS) Facility, which is owned and operated by Wright State University (WSU), but housed in Building 441 at Wright-Patterson AFB (WPAFB), was substantially upgraded, replacing aging and unreliable hardware and providing enhanced capability, including brighter and higher resolution projection and increased computational speed. The CAVE® is a 10-ft by 10-ft by 10-ft cubical display system (with four rear projection screens as walls and a top projected floor). The upgrade included 5 new DLP Barco Galaxy NW-12 projectors, new mirrors, new video cabling, new stereoscopic emitters, and a new Barco Xracku touch screen remote control system for the CAVE®. In addition, the CAVE® floor was re-painted and the InterSense IS-900 tracking system was re-calibrated. AFRL also arranged to have the interior of the chamber housing the CAVE® painted black reducing light reflections. In

parallel, we installed 5 new Boxx Technologies VisBoxx computers replacing the old image generation cluster. A number of enhancements, bug fixes, and experiment specific modifications to the software have been implemented as part of this contract. Some of this work was performed by VERITAS staff and other aspects were completed through a subcontract with Middendorf Scientific Services, Inc. Because of time and budget constraints, in the initial implementation of the IVP, auditory communications between VERITAS and AVL relied on phone lines. This system proved to be both noisy and unreliable. To remedy these problems a new VoIP-based system (teamspeak) was selected, implemented, and evaluated. Teamspeak was in turn integrated with the HLA SLAB application, which spatializes the voice communication and includes push-to-talk capability. The software system we developed to interact with the subjects and control the visualization, the Integrated Visual Environment (IVE), was modified to support rendering and evaluation of elevated auditory sources. In addition, the crude helicopter motion model that had been implemented for the demonstration version of the IVE was enhanced to allow manual adjustment of the automated flight path, providing a mechanism to increase the proportion of time that the ground-based subject avatar was visible to the helicopter controller. The helicopter motion model now orbits the location of the ground-based user (with a changeable radius and altitude) rather than orbiting the entire city.

4.0 METHODS

4.1. Subjects

A total of 12 subjects (10 males and 2 females) participated in a "search-and-rescue" scenario. The subjects were paired into 6 teams, with one subject acting as a ground-based “PJ” and the other acting as a helicopter-based “Spotter.” The subjects ranged in age from 19 to 30 years old.

4.2. Task

The PJ’s task was to move through a 100-m by 100-m simulated urban environment in order to find a downed pilot. Along the way, the PJ could encounter enemy (eight) and civilian (eight) personnel (only distinguishable based on whether their weapon was raised); PJs were instructed to shoot (by pointing a handheld “wand”) the enemies before they themselves where shot, but to avoid shooting civilians.

The Spotter viewed the same urban environment and could see (except when obstructed by buildings) the enemy and civilian personnel and an avatar representing the PJ. The Spotter’s task was to guide the PJ to the downed pilot as quickly as possible and indicate the presence and location of enemies.

The PJ interacted with the simulated urban environment while seated in the VERITAS CAVE® at WPAFB and the Spotter interacted with the environment while seated in a CAVE-like facility with three walls and a floor (the iSpace) that is housed at WSU and managed by daytaOhio.

4.3. Conditions

A total of nine conditions were formed by combining each of three PJ Display conditions with each of three Spotter Display conditions. In the Visual PJ Display condition the PJ received

communications from the Spotter via a chest mounted visual display; there was no auditory comm channel. In the Mono Audio PJ Display condition, these communications where delivered verbally via a monaural headphone system. In the 3D Audio PJ Display condition, these communications were "spatialized" and displayed through headphones so that the PJ heard the Spotter's voice as arising from a location of the Spotter's choosing within the simulated urban environment in which the PJ was working. The Spotter Display conditions differed from one another in terms of the manner in which the Spotter received information about the urban environment and the way in which the spotter marked positional information in the PJ's environment. In the Laser Spotter Display Condition, the Spotter viewed the urban environment from the side door of a simulated helicopter that circled above the PJ and used a simulated laser range finder to mark locations on the ground below (to present positional information to the PJ except in the Mono Audio PJ Display condition); "pulling the trigger" on the range finder also opened the audio comm channel to the PJ (except in the in Visual PJ Display condition). In the Tablet Spotter Display condition, the Spotter viewed the urban environment from the side door of the helicopter, but marked positional information on a map of the urban environment presented on a "tablet" computer (for convenience a laptop computer and mouse was used to approximate a wearable or tablet computer that might be used in the field). The locations of the PJ and Pilot were shown as symbols on the map display. Again, depending on the PJ Display condition, marking a location presented positional information to the PJ and opened the comm channel. In the God's Eye Spotter Display condition, the spotter sat in the iSpace with the projectors off and viewed the simulated environment only via a map displayed on the tablet computer. In this case, however, the location and movements of all personnel were indicated on the map. In addition, enemy and civilian personnel were clearly indicated by color. Thus, this display (see Figure 2) provided a detailed view of the entire environment (the God's Eye condition). Again, depending on the PJ Display condition, marking a location presented positional information to the PJ and opened the comm channel.

Under the 3D audio condition, communications were spatialized using the SoundLAB (SLAB) software developed at NASA (Wenzel, Miller, & Abel, 2000). Individualized Head-Related Transfer Functions (HRTFs) were recorded for 5 of the 6 subjects with the role of PJ. For the 6th subject, individualized HRTFs were not available and one of the other subject's HRTFs were used.

4.4. Design

Each pair of subjects completed all nine (3 PJ Display conditions* 3 Spotter Display conditions) treatment combinations in sets of nine trials each, with a single combination of PJ Display and Spotter Display being used throughout a set. Each trial within a set consisted of one trip through an urban environment from the start point to the downed pilot. Each of the nine trials within a set was conducted through a different randomly generated urban environment, created from the combination of nine maps with three initial positions (i.e., PJ location and downed pilot location) on each map. In determining the order in which conditions were presented to the subjects, we blocked the design by

Spotter Display, so that Spotter Display level was held constant while all three PJ Display conditions were run. The six possible orders of PJ Display condition by the six possible orders of Spotter Display condition yielded 36 possible orders of treatment combinations. So, the order of the PJ display was further confounded with the order of the Spotter Displays using mirrored Latin squares, so that only 6 order combinations appeared in the experiment. As is the case in all cases of fractionation, higher order interactions of position variables constitute possible, though non-parsimonious explanations for treatment effects.

5.0 RESULTS AND DISCUSSION

A number of dependent measures were examined. Means and standard errors are shown in Tables 1-6 for Time to Completion (elapsed time from the start of the trial until the pilot was found), Distance Traveled (the distance the PJ traveled from the starting point to the downed pilot), Number of Times Shot (the number of times enemy personnel shot the PJ -- the PJ was not “killed” when shot and the trials continued until the Pilot was found), the Number of Enemies that Shot at the PJ, the Number of Enemies Shot by the PJ, and the Number of Civilians Shot by the PJ, respectively.

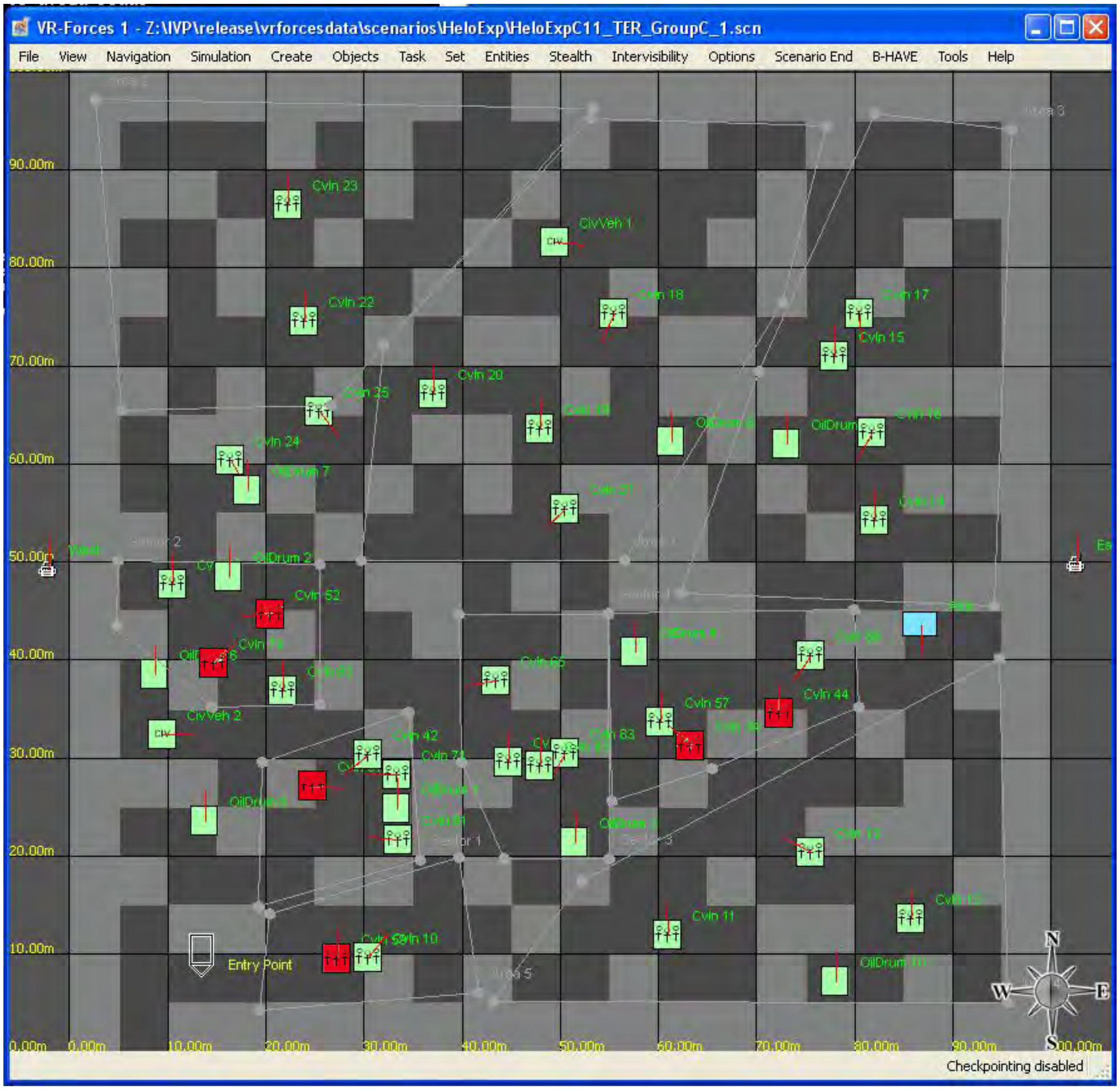

Figure 2. Screen shot of a display comparable to the one used in the God's Eye Spotter Display condition. Dark squares are streets and light squares are buildings. Red markers indicate the locations of enemy personnel. Green and blue markers indicate civilians and friendly forces.

A 3 X 3 randomized blocks ANOVA was computed for each dependent measure. Only two F scores were found to be significant. As shown in Figure 3, there was a significant main effect of Spotter Display on the Number of Enemies that Shot at the PJ ($F(2,10)=7.13$, $p=.01$), but neither the main effect of PJ Display ($F(2,10)= 0.59$, $p=.57$) nor the interaction between PJ Display and Spotter Display ($F(2,10)=0.71$, $p=.59$) were significant. Post hoc Tukey HSD tests indicate that fewer enemies shot at the PJ when the Spotter used the God's Eye Display as compared to the Tablet Display ($p<.01$). This result is not surprising given that more information was immediately available to the Spotter in the God's Eye Spotter Display condition; and so, presumably, the Spotter was better able to help the PJ minimize contact with the enemies. Post hoc Tukey HSD tests indicated that the

God's Eye Display and the Laser Display did not differ significantly (p=.08) nor did the Laser display and the Tablet display (p.=55).

Table 1. Completion Time (s) in each of the nine conditions.

	Spotter Display Condition					
	Laser		Tablet		God's Eye	
PJ Display	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual	88.6	25.5	83.9	11.6	90.1	18.5
Mono	91.9	20.4	85.8	11.6	87.6	12.1
3D Audio	92.9	21.7	88.9	18.9	90.1	13.1

Table 2. Distance Traveled (m) in each of the nine conditions.

	Spotter Display Condition					
	Laser		Tablet		God's Eye	
PJ Display	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual	161.6	20.6	154.7	14.4	151.5	16.3
Mono	163.2	21.7	160.8	17.7	158.5	16.9
3D Audio	162.6	19.3	160.3	20.0	161.8	17.2

Table 3. Number of Times Shot in each of the nine conditions.

	Spotter Display Condition					
	Laser		Tablet		God's Eye	
PJ Display	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual	12.7	8.9	13.5	7.8	12.5	8.8
Mono	11.9	7.5	10.6	6.2	9.2	7.2
3D Audio	11.6	6.7	11.5	6.4	10.8	7.6

Table 4. Number of Enemies That Shot at PJ in each of the nine conditions.

	Spotter Display Condition					
	Laser		Tablet		God's Eye	
PJ Display	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual	4.9	1.6	5.0	1.7	4.6	1.8
Mono	4.6	1.9	5.0	1.7	4.0	1.6
3D Audio	4.8	1.7	4.9	1.7	4.5	1.8

Table 5. Number of Enemies Shot in each of the nine conditions.

	Spotter Display Condition					
	Laser		Tablet		God's Eye	
PJ Display	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual	4.0	2.5	4.3	1.8	5.0	1.9
Mono	4.6	2.1	4.9	1.8	5.0	1.6
3D Audio	4.5	2.0	4.9	1.7	4.7	1.7

Table 6. Number of Civilians Shot in each of the nine conditions.

	Spotter Display Condition					
	Laser		Tablet		God's Eye	
PJ Display	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual	0.8	1.0	0.9	1.1	0.9	1.9
Mono	1.0	1.3	0.8	1.4	1.0	1.4
3D Audio	0.7	1.4	0.6	0.9	0.9	1.0

As shown in Figure 4, there was a significant main effect of PJ Display on the Number of Enemies Shot by the PJ ($F(2,10)=4.20$, $p=.05$), but neither the main effect of Spotter Display ($F(2,10)= 0.54$, $p=.60$) nor the interaction between PJ Display and Spotter Display (($F(2,10)=1.09$, $p=.59$) were significant. Post hoc Tukey HSD tests indicate that a larger number of enemies were shot under the Mono Audio PJ Display condition as compared to the Visual PJ Display condition ($p=.04$). This finding was expected because the visual display was a "heads down" display, and so, it should have been more difficult to quickly locate and address enemy threats. It was at least somewhat surprising that the 3D Audio PJ Display condition did not show a similar advantage when compared to the Visual PJ Display condition ($p=.17$). The Number of Enemies Shot under the Mono Audio PJ display condition and the 3D Audio PJ Display condition did not differ significantly ($p=.77$), even though the more immediate spatial information available in the 3D Audio PJ Display condition was expected to enhance situation awareness and speed responding.

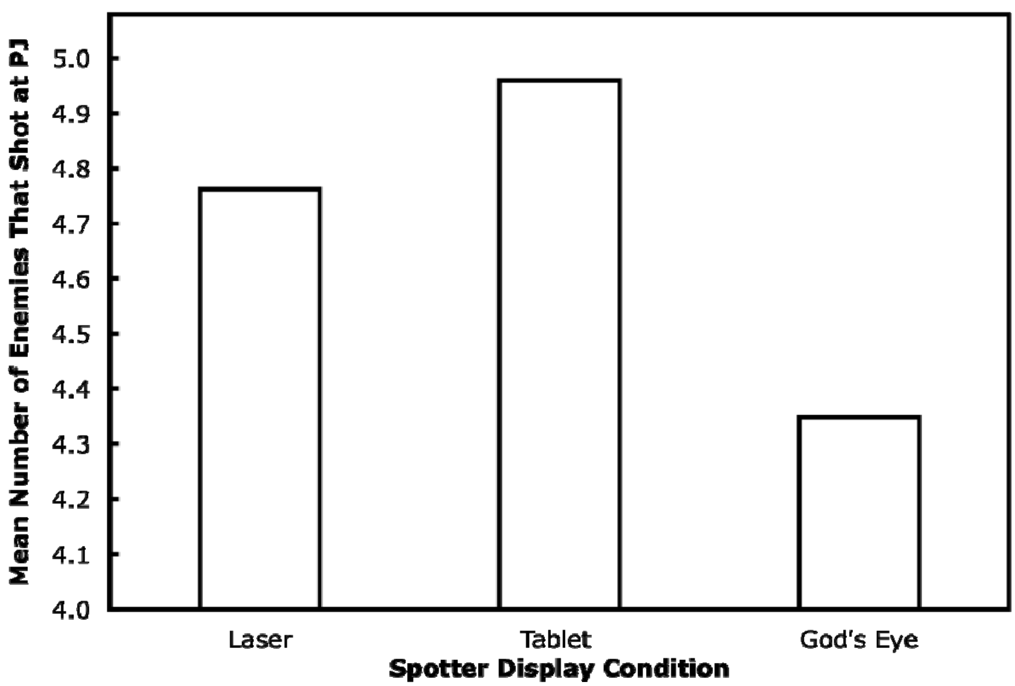

Figure 3. Mean Number of Entities that Shot at the PJ under the three Spotter Display conditions.

Overall, the results did not show any major differences in performance across conditions. Even the differences that were found to be statistically significant were relatively small. Nevertheless, these results are important both as a demonstration of capability and by providing a number of lessons learned which are being addressed in subsequent research and development. In order to perform this research a rather complex distributed virtual environment was developed by a relatively

small team of scientists and engineers. The team employed custom (in house), open source, and COTS elements to develop a 3D urban environment populated by entities (enemies and civilians controlled by artificial intelligence subroutines), which allowed realtime interaction (auditory and visual), between two users working in facilities at different institutions, and between the users and artificial entities. The experiment was executed without major system difficulties. This distributed virtual environment facility (the IVP) is now supporting additional research on display design and will continue to provide a complex and flexible simulation environment to support future work in a variety of domains.

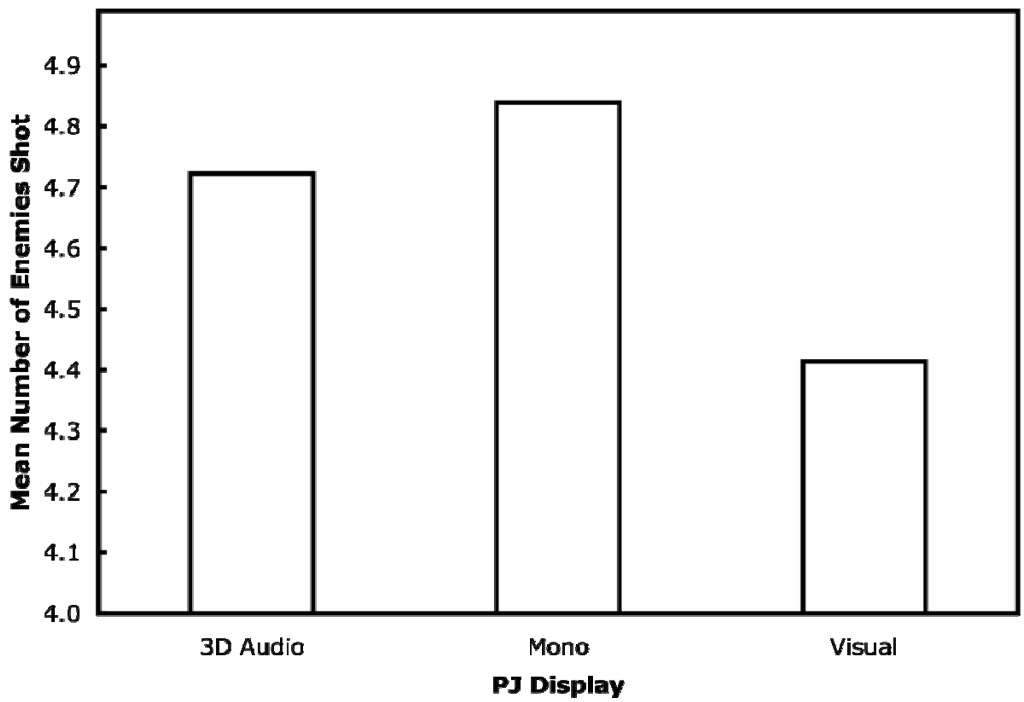

Figure 4. Mean Number of Enemies Shot by the PJ under the three PJ Display conditions.

Informal trials (with scientists and engineers working on the project acting in the roles of PJ and Spotter) indicate that the small size of the observed differences resulted, at least in part, because of ceiling effects. The scenarios for this study were not particularly difficult and may not have created significant cognitive demand on the team members. The virtual city was small (100 m by 100 m). In addition, there was a relatively small number of hostile entities that "fought in the open." Therefore, the PJ's task could be completed effectively with little information from the Spotter. Moreover, because neither the PJ nor the Spotter were particularly taxed by their tasks, the manner in which the information was generated and displayed may have had little impact on the PJ's ability to utilize the information.

Because of these ceiling effects it is difficult to evaluate whether audio annotation will provide significant benefit in an operational setting. We anticipate that such benefits will be more readily demonstrated in environments in which it is more difficult to find and identify targets, where the cognitive demands on the PJ and spotter are greater, and/or when spatial information needs to be provided to multiple observers simultaneously. These issues will be examined in future research projects.

6.0 References

Ephrem, A. J., Finomore, V. S., Gilkey, R. H., Newton, R. M., Romigh, G. D., Simpson, B. D., Brungart, D. S., & Cowgill, J. L. (2008). An "Audio Annotation" technique for intuitive communication in spatial tasks. In Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting, 52, 2127-2131.

Wenzel, E. M., Miller, J. D., & Abel, J. S. (2000). Sound Lab: A real-time, software based system for the study of spatial hearing. Proceedings of the 108th Convention of the Audio Engineering Society, Paris, Feb. 2000, New York: Audio Engineering Society, Preprint 5140.